

Microsoft Enterprise Consortium: A Resource for Teaching Data Warehouse, Business Intelligence and Database Management Systems

Jennifer Kreie
jkreie@nmsu.edu
Accounting & Information Systems
New Mexico State University
Las Cruces, New Mexico 88003, United States

Shohreh Hashemi
hashemis@uhd.edu
Finance, Accounting & Computer Information Systems
University of Houston-Downtown
Houston, Texas, 77002, United States

Abstract

Data is a vital resource for businesses; therefore, it is important for businesses to manage and use their data effectively. Because of this, businesses value college graduates with an understanding of and hands-on experience working with databases, data warehouses and data analysis theories and tools. Faculty in many business disciplines try to prepare students for this data-oriented business environment by teaching database and business intelligence concepts in their courses. However, many faculty may not be aware of various important online resources. For example, the Microsoft Enterprise Consortium (MEC) provides data and instructional resources to faculty in business colleges through access to several data sets that could be used for teaching a range of topics including fundamentals of database management, beginning and advanced SQL, data warehouses, and business intelligence (data mining, data cubes and dimensional reporting). The purpose of this paper is to introduce the Microsoft Enterprise Consortium, to discuss the various teaching resources it provides, to present an overview of the materials available through the MEC, and to give an example of successful implementation and utilization of MEC resources.

Keywords: data warehouse, business intelligence, database, data cube, dimensional reporting, data mining, teaching resources

1. INTRODUCTION

Businesses aim to maximize their data resource (Preston, R., 2010; Ferguson, R., 2008; Britt, P., 2006) and they value employees with data-related skills (Downey, McMurtrey and

Zeltmann, 2008). Business faculty in many disciplines recognize that their students need to work with data and be familiar with software tools used to manipulate and analyze data (Jukic, N. and Gray, P., 2008). A comprehensive valuable data-related resource is available to

faculty through the Microsoft Enterprise Consortium (MEC). It is worth noting that there is another data-oriented teaching resource called the Teradata University Network (Jukic, et al, 2008; Winter, R., Gericke, A., and Bucher, T., 2008). Both the MEC and the Teradata sites are valuable resources for teachers and students.

The MEC site in particular contains real-world data sets that are excellent for incorporating into business intelligence and database coursework. For fundamental database concepts, the MEC provides databases that range from fairly simple in terms of the number of tables, relationships and data rows to fairly complex with a large volume of data. For teaching data warehouses and business intelligence, the MEC provides very large data sets contributed by companies such as Sam's Club, Tyson Foods, and Dillard's. These companies have made "real" data available for instructional purposes through the MEC.

The MEC is hosted at the University of Arkansas Walton College of Business (<http://enterprise.waltoncollege.uark.edu>). Although the resources are free, faculty must register themselves and then add their students through a course management interface. The data sets are maintained in SQL Server 2008 and access is through remote desktop software.

The resources are organized by topic area and many topic areas include PowerPoint slides (Figure 1 in Appendix A), videos, assignment problems and solutions. The instructor-only material, such as instructor guides, assignments and solutions, will soon be housed in a password-protected section of the Microsoft Faculty Connection web site (<http://www.microsoft.com/education/facultyconnection>). Faculty who use the MEC site in their classes are encouraged to expand the reserve of teaching materials by contributing additional presentations and assignments.

The MEC is a valuable resource for IS educators because its utilization in IS courses facilitates teaching and provides students with an array of hands-on exercises using "real" and large data sets.

In what follows, there is an overview of the MEC site's available and soon-to-be-released data sets, instructional materials, and examples of how such instructional material could be utilized in various courses are presented.

2. Databases at the MEC

The MEC offers several data sets to educators. In addition to several small-scale databases, Dillard's Department Stores, Sam's Club and Tyson Frozen Foods have donated large data sets. These large data sets are particularly useful when covering topics in SQL tuning, data warehouses, business intelligence and data mining.

Microsoft's AdventureWorks

Microsoft's AdventureWorks is a sample data set that includes a transactional database and a data warehouse for analysis purposes. This database has numerous tables and provides examples of data in several business areas:

- Sales and marketing
- Product/inventory
- Purchasing
- Manufacturing
- Human resources

This sample data is based on a fictional bicycle manufacturer. (Note: There is an AdventureWorksLT database that is also available at the MEC. Though it is a smaller and simpler data set, it is denormalized.)

Hallux Productions

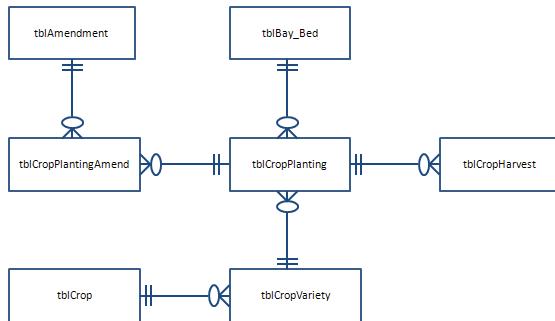
This data set is based on a music publishing and management company scenario (Hallux). There are 25 tables. This business scenario includes the sales order data model so often depicted in textbooks; however, it extends this model and tracks data about bands, band members, recordings, performances and more. The data spans several years and provides numerous ways to explore data modeling concepts and teach simple to complex SQL statements. See Figure 2 in Appendix A for the data model.

Greenhouse

The greenhouse database is based on a real greenhouse operation near New York City that is part of a non-profit farm and sustainable living education center. Like the Hallux Productions database, this data set provides teachers and students with a data model not typically found in textbooks. The database tracks information about different zones within a greenhouse, the crops planted, and the amendments used in the soil and the crop harvests. Figure 3 shows the

greenhouse data model. This is a simple data model that could be used when teaching fundamental and advanced SQL. Some tables have only a few rows of data and others, such as harvests, have a few thousand rows.

Figure 3. Greenhouse ERD

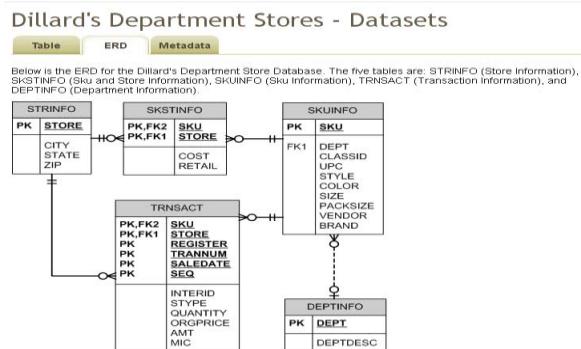


Dillard's Department Stores

The Dillard's Department data set contains sales transaction data for 453 stores during 2004 and 2005 (Dillard's). This database has several tables, one of which has over 120 million rows of data.

Figure 4 shows the Dillard's data model. This figure also shows the standard information about the large data sets provided at the MEC site. There is a general description of the business, some example table data, an ERD and metadata.

Figure 4. Dillard's ERD



Sam's Club

This is a data set of retail sales data. There are six tables related to sales, one of which has more than 48 million rows. The tables contain data from one month of transactions. The MEC

server also has modified versions of the Sam's Club data set that are meant specifically for teaching data warehouses and data mining.

Tyson Frozen Foods

The Tyson data is a data cube with two years of sales transactions in a fact table linked to several dimension tables, including business division, pricing segment, and product. There are over 11 million rows of data. Figure 5 in Appendix A shows the Tyson data model.

3. Instructional Materials at the MEC

Each topic area has a set of videos and PowerPoint slides and most include exercises for students and solutions to the exercises. These materials were designed to give a short presentation on a specific topic. Each presentation is kept brief — 10 to 15 slides usually — and the videos usually run under 10 minutes. An instructor might choose to use the entire set of presentations within a topic area or take the one or two he/she finds useful.

Database Concepts

This subject area contains materials that cover the subjects listed below and includes a "test your knowledge" exercise that covers relational database fundamentals. Currently, the topics covered are:

- Data Models
- Relationships
- Logical & physical models
- Primary & foreign keys
- One-to-many, one-to-one, and many-to-many relationships
- Multivalued attributes

The videos for this subject area are also available on YouTube. A search for "MEC database" will list these videos.

SQL Fundamentals

The SQL fundamentals module uses two databases. A student-teams database is created and populated in the first couple of presentations, and it along with the Microsoft AdventureWorks2008 database are used in subsequent presentations. The SQL fundamentals include the following topics:

- Creating tables and populating them (SQL scripts included)
- Introduction to SELECT ... FROM ... WHERE
- Logical operators in the WHERE clause
- Using inexact criteria and ranges of values (inexact matching)
- IN and NOT IN
- NULL and NOT NULL
- Joining multiple tables either with JOIN or in the WHERE clause
- Organizing output: sorting, column aliases and dynamic columns
- Self-joins and table aliases
- Aggregate queries
- Traditional set operators

SQL Advanced Features

The advanced SQL materials use AdventureWorks2008, student-teams, and the greenhouse databases. Advanced SQL covers these topics:

- One-sided outer join and full outer join.
- One-sided outer join and a NULL criterion
- Nested queries: Type I & Type II
- In-line queries in the FROM clause
- Delete from a table using a subquery
- Difference problems
- Nested aggregates
- Division problems

Business Intelligence

The Microsoft Analysis Services software is used in this subject area. The instructional material covers:

- Building a data cube (one for Sam's Club and for Dillard's data)
- Using a pre-built cube (Dillard's data)
- Dimensional reporting using Microsoft Reporting Services (Sam's Club data)
- Data mining introduction
- Data mining with a decision tree
- Data mining with neural networks

4. Classroom Examples

The instructional materials at the MEC can be scaled and used as supplementary materials in introductory to advanced Information Systems courses, as well as in Supply Chain, Marketing and other business courses. For use in introduction to data bases courses, the EMC instructional materials can be scaled down to explain databases, tables, rows, etc., and to

introduce data warehouses, data mining, and business intelligence concepts. At the other end of the spectrum, the instructional materials at the EMC can be fully utilized to build cubes, manipulate and analyze data and trend to devise strategic plans of action.

In fact, the MEC instructional materials can be easily aligned with a typical database textbook to complement coverage of various topics. For example, when teaching introductory database concepts—data models, database design concepts, relational database models, entity relationship modeling, normalizations of database tables, and advanced data modeling—a walk-through of MEC tables could be conducted and some or all of the instructional materials could be utilized. This practice would provide students a rich and “real” set of resources to draw upon when studying topics such as rows, columns, and populating tables. Later, when advanced database design and implementation concepts such as SQL, database performance tuning and query optimization topics are covered, more advanced topics covered in MEC instructional modules such as SQL advanced materials could be utilized, followed by a series of hands-on exercises. Finally, the MEC's Microsoft Analysis Services instructional materials could be used to introduce data warehouses, data mining and business intelligence, also followed by a hands-on practice when students would create data cubes, analyze data, identify trends, and devise strategic plans. In what follows, specific examples of possible uses of the EMC data models are presented.

Pre-Built Data Cube

In a long semester (15 weeks duration) database class, two to four class sessions could be set aside to cover data warehouse and cube concepts, and another two or to four sessions could be devoted to hands-on database manipulation exercises. Because of the short time frame, the instructor would use the pre-built cube from the Dillard's data set. In addition, to augment the learning process, one could use the MEC handouts that introduce students to Microsoft Analysis Services and use of the Dillard's cube to analyze the data. One teaching scenario could be that after going through the MEC handouts, students are assigned a set of questions that would require use of the data cube to obtain answers. Another teaching scenario could be that, once coverage of the materials on the MEC handout is

completed, students are assigned open-ended questions that would require them to explore the pre-built cubes to find significant trends and present their findings in a short write-up. These types of exercises provide students with hands-on experience working with pre-built cubes while enhancing their critical thinking skills and effective business intelligence analysis.

SQL & Query Design

In a database class that spends several weeks covering SQL, the very large data sets can illustrate query performance. This is an advantage over small data sets, such as those typically provided with textbooks, where there's no discernible difference between "good" and "bad" queries because all queries run very quickly. As an example, the queries shown in Figures 6 and 7 have the same output, but their structure could be the basis of discussion about SQL query design using a subquery or "distinct."

Figure 6. Example 1 SQL

```
select distinct skuinfo.dept
, deptdesc
, vendor
, skuinfo.sku
from ua_dillards.dbo.deptinfo
join ua_dillards.dbo.skuinfo
on deptinfo.dept = skuinfo.dept
join ua_dillards.dbo.trnsact
on skuinfo.sku = trnsact.sku
where saledate between '2005-08-01' and '2005-08-07'
order by skuinfo.dept, vendor;
```

Figure 7. Example 2 SQL

```
select skuinfo.dept
, deptdesc
, vendor
, sku
from ua_dillards.dbo.deptinfo
join ua_dillards.dbo.skuinfo
on deptinfo.dept = skuinfo.dept
where sku IN
(select sku
from ua_dillards.dbo.trnsact
where saledate between '2005-08-01' and '2005-08-07'
)
order by skuinfo.dept, vendor;
```

There's another opportunity for teaching based on the two queries above due to the fact that they result in 253,715 rows of output. Students used to small databases and inclined toward lazy coding, such as `SELECT *` and no filtering in the `WHERE` clause, can learn from a large data set

that careful planning and design are needed to make output practical and useful. Below is a modified query that reduces the output from over 253 thousand rows to 13 rows. The modified query shown in Figure 8 simplifies output by (1) counting the SKU items sold by vendor, rather than listing each sale, (2) showing sales in a particular department (ESPRIT) and (3) showing only the two stores in Tallahassee, Florida (stores #4302 and 4502).

Figure 8. Modified Query

```
/*
The WHERE clauses limits items sold to those in
the ESPRIT department in the Dillards
stores in Tallahassee, Florida (store # 4302, 4502).
*/
select skuinfo.dept
, deptdesc
, vendor
, count(sku) as "SKU count"
from ua_dillards.dbo.deptinfo
join ua_dillards.dbo.skuinfo
on deptinfo.dept = skuinfo.dept
where deptdesc = 'ESPRIT'
and sku IN
(select sku
from ua_dillards.dbo.trnsact
where saledate between '2005-08-01' and '2005-08-07'
and store in (4302, 4502)
)
group by skuinfo.dept, deptdesc, vendor
order by skuinfo.dept, vendor;
```

5. Summary

This paper has presented an overview of the instructional resources and data sets available at the Microsoft Enterprise Consortium. These resources cover a range of important topics that most business students and, certainly, students majoring in information systems should learn. The MEC is a free resource to higher education; however, faculty must register themselves and their students to gain access to its resources. Faculty who join the MEC are invited to compile and submit to MEC new course material they develop using these databases for possible inclusion in the MEC resources. A community of faculty could help the MEC continue to expand teaching materials provided to teachers and students.

6. REFERENCES

Britt, P. (2006). Trends in Database Management Systems Technology. *InformationToday*, 23(5), pp. 19-20. (Magazine Article)

Dillard's Data Model. Retrieved May 11, 2011 from

<http://enterprise.waltoncollege.uark.edu/1683.asp>. (Web page)

Downey, J. P., McMurtrey, M. E. and Zeltmann, S. M. (2008). Mapping the MIS curriculum based on critical skills of new graduates: an empirical examination of IT professionals. *Journal of Information Systems Education*, 19(3), pp. 351-364. (Journal 3 or more Authors)

Ferguson, R. (2008). BI on the move. *eWeek*, 25(11), pp. 8-11. Retrieved from EBSCOhost. (Magazine Article)

Hallux Productions Data Model. Retrieved May 12, 2011 from <http://enterprise.waltoncollege.uark.edu/1684.asp>. (Web page)

Jukic, N. and Gray, P. (2008). Teradata University network: a no cost web-portal for teaching database, data warehousing, and data-related subjects. *Journal of Information Systems Education*, 19(4), pp. 395-403. (Journal 2 Authors)

Microsoft AdventureWorks Complete Model, Retrieved May 12, 2011 from <http://msdn.microsoft.com/en-us/library/bb399790%28v=vs.90%29.aspx>. (Web page)

Microsoft Enterprise Consortium, <http://enterprise.waltoncollege.uark.edu/>. (Web page)

Microsoft Faculty Connection, <http://www.microsoft.com/education/facultyconnection>. (Web page)

Preston, Rob. (2010). Analytics for Every Action. *InformationWeek*, May 24. P. 48 (Magazine Article).

Tyson Frozen Foods Data Model. Retrieved May 12, 2011 from <http://enterprise.waltoncollege.uark.edu/1685.asp>. (Web page)

Winter, R., Gericke, A., and Bucher, T. (2008). Using Teradata University Network (TUN), a free internet resource for teaching and learning. *Educational Technology & Society*, 11(4), pp. 113-127. (Journal 3 Authors)

Appendix A

Figure 1. Example of PowerPoint slides
(6 of 11 slides are shown)

Slide 1: SQL Fundamentals

SQL Fundamentals
SELECT ... FROM ...

Microsoft Enterprise Consortium: <http://enterprise.waltoncollege.uark.edu>
Microsoft Faculty Connection/Faculty Resource Center <http://www.facultyresourcecenter.com>

Slide 2: What you'll need ...

- For this and other SQL lessons, you need a **user account** from the **Microsoft Enterprise Consortium**. Get this account from your instructor.
- Log in to MEC for this lesson and into MSSMS (Microsoft SQL Server Management Studio).
 - Be sure to select your account ID under Database in the Object Explorer pane, similar to the example shown here.

Slide 3: AdventureWorks

Microsoft provides an example database called AdventureWorks (AW). For the time being, only a subset of tables from this database will be used.

Shown here is the data model for the HR portion of the AW database. An additional table is included from the PERSON section of the database.

Slide 4: Student-Teams

This database keeps information about students, the teams they are assigned to and the peer evaluations students complete for their teammates at the end of a project.

Slide 5: SELECT ... FROM

- The SELECT statement has several components but let's start with only the FROM clause.
- A simple SELECT statement has the following format ...


```
SELECT <column1>, <column2>, ...
  FROM <tablename>;
```
- The angle brackets indicate words that change depending on what data we want to see.

Recommendation: Keep all the SQL commands you write for a lesson in a text file using Notepad. Do the same for SQL you write or assignments.

Slide 6: Query the Student-Team database

- Open a Query pane (New Query).
- Type the SQL in Notepad then copy/paste into the query pane. Add a comment.
- Let's start off by seeing what teams there are.
- Enter two columns (teamID and team_name) in the SELECT clause.
- List the table name in the FROM clause.
- Execute the query.

```
/* Show teams */
select teamid, team_name
from teams;
```

teamid	team_name
1	DBMS Associates
2	EDSIG, Inc.
3	DIGITAL
4	ITPROS
5	ITRON
6	SYNTEL
7	SYDES
8	TECHS

Figure 2. Hallux ERD

Hallux Productions - Datasets

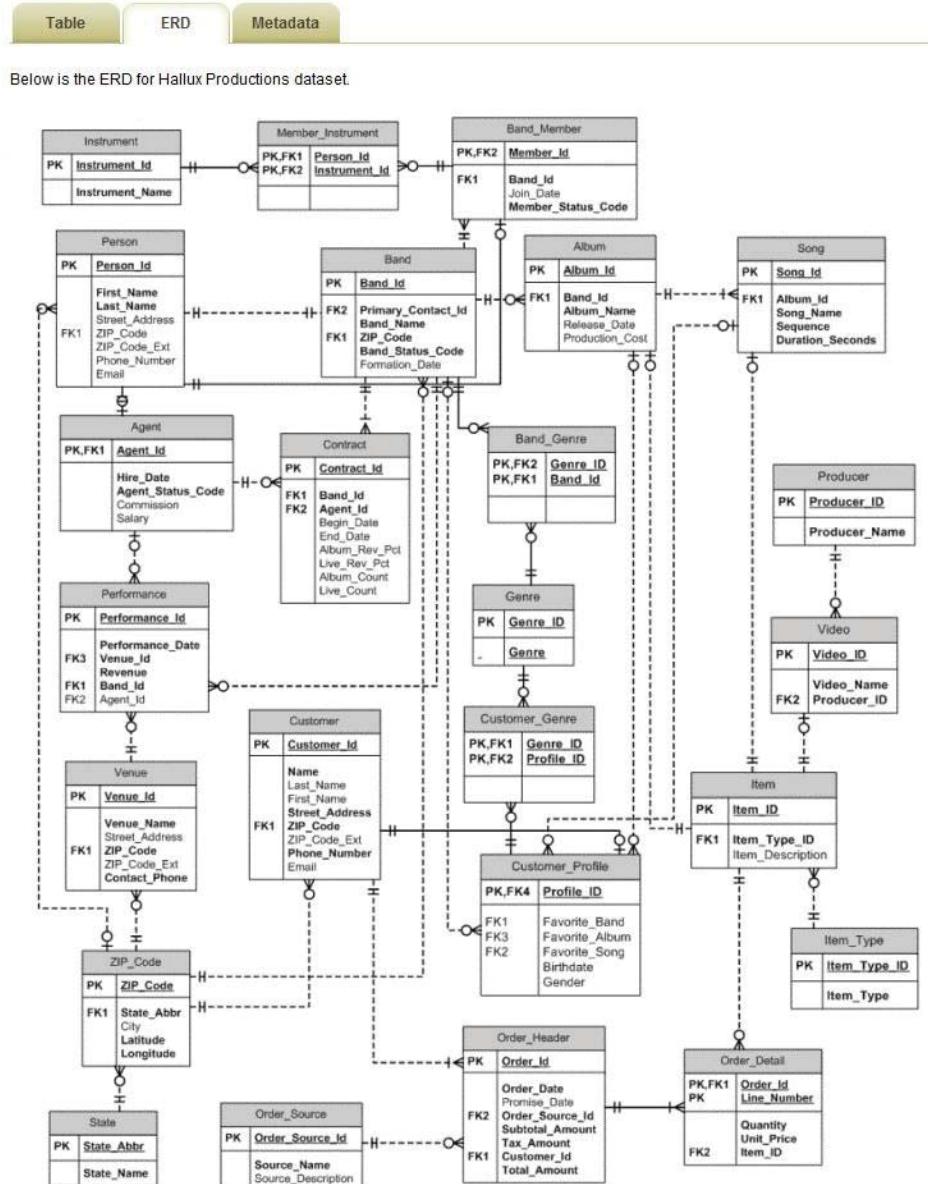


Figure 5. Tyson Frozen Foods ERD

Tyson Frozen Foods - Datasets

